系統樹の可視化：
木構造の情報可視化における色とレイアウトの統合アプローチ

東京科学大学　修士1年　田中愛万音

はじめに

現代社会では、データの急増に伴い、その有益な情報を引き出し理解することがますます困難になっている。特に、データが階層的な構造を持つ場合、その複雑さを視覚的に理解するための効果的な情報可視化手法が求められている。過去数十年に渡り、様々な可視化手法が提案され、研究が進展してきた。例えば、Treemaps、Icicle Plots、Sunburst 図など、多くの静的可視化手法が数多くの分野で利用されている。しかし、これらの手法は主に平坦なデータ構造を想定しており、大規模で複雑な木構造に焦点を当てたものは限られている。本研究では、大規模で複雑な木構造データの可視化に焦点を当て、その課題と解決策を探求する。

大規模木構造の課題

(I-1)認知的負荷

ノード数が多い・階層が深い → どこに焦点を当てればよいか分からない・階層構造を理解しにくい
ユーザーは常に全てのデータを求めてはいない

(I-2)色の効果的な適用

色は順序性やトポロジカルな関係を表現するために利用されるが、大規模な木構造では色の数が増えるとともに、その効果的な適用が難しくなる。異なる階層や部分木において色差を維持することが困難である。

設計

上述の2つの課題に対してそれぞれ解決策を講じる。

(S-1)認知的負荷の対処案

ユーザーは常に全てのデータを求めていないということに留意し、インタラクションを使用して、ユーザーが求めている情報だけに焦点を当てることにする。このようなインタラクションはフィルターなど様々なものが挙げられるが、本研究では**ズーム機能**に着目した。

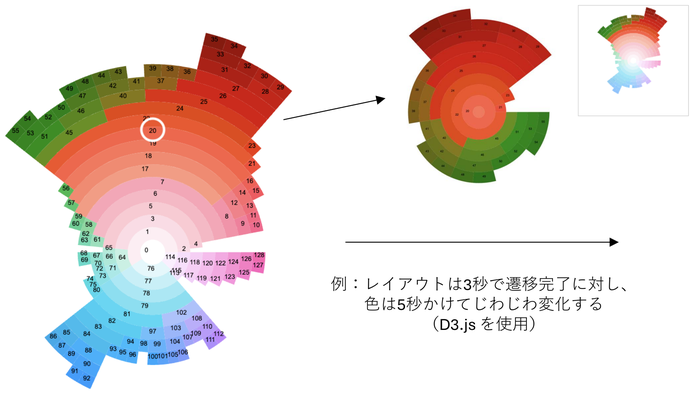
(S-2)色の効果的な適用の対処案

ズーム操作により、色空間で変換を行い、**色を動的に変化**させる。ユーザーがズームをした際にその周辺のノードを識別できる一定の色差を設ける。
人間が色の変化に気づかない方が分析に良い → 色変化のスピードを慎重に設定する

提案手法

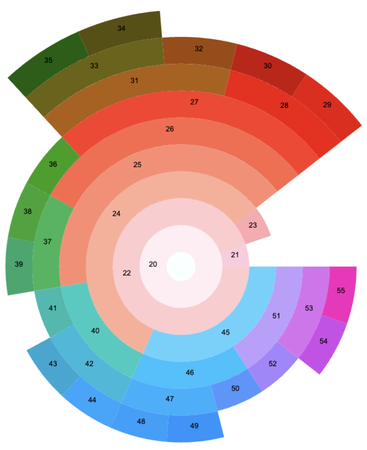
課題を考慮した解決策を反映した可視化手法を提案する。
ユーザーが特定のノードをクリックすると、従来はレイアウトが変化するのみだったのに対し、新しいズーム機能では、レイアウトと色が同時に変化し始め、色の方が遅く変化する。
従来の Plotly のグラフでは、ズームをした際の遷移後グラフの表示では元のグラフの部分が見えなくなり情報を得ることが不可能になる。しかし提案手法では、クリック遷移後のグラフに表示されるノードの透明度による強調により、従来ではグラフ遷移後に見えなくなっていたノードを含む Overview により、Context を維持しながら拡大することができる。

従来の遷移後グラフ
（Plotly.Express）



例：レイアウトは3秒で遷移完了に対し、
色は5秒かけてじわじわ変化する
（D3.js を使用）

新しいズーム機能



実装

提案手法の実装において、新しく使用する D3.js の特徴と、扱う実データについて述べる。

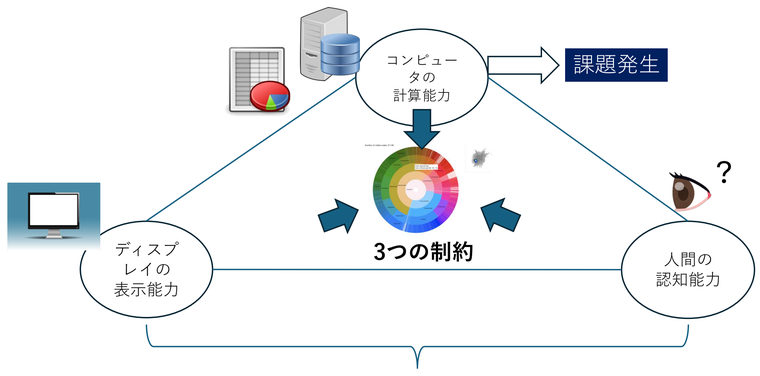
- **D3.js** について
  従来はグラフの可視化ツールとしてデフォルトの設定が充実している Plotly を使用していた。一方、提案手法では、柔軟なデザインやインタラクションの制御が必要なため、D3.jsの持つ自由度が特に有用である。よって本研究の実装では、D3.jsを使用する。

- データセットについて
  可視化で扱うデータは、「Catalog of Life」という系統樹のデータである[COL]。これは種族名、個体名も含め全部で約250万ノードである。これにはもちろんヒト（Homo sapiens）も含まれており、右はドメイン（真核生物）から種（ヒト）まで、ヒトがどのような分類階層に属しているかを階層的に示している。元のデータセットは、ID、parentID、scientificName、rankの4つのカラムを含む。parentIDは、各生物の親（上位分類）の識別子である。今回は「ID」カラムと「parentID」カラムを用いて、各生物の親子関係を可視化する。
  事前にデータについて分析したところ、最大深さのノードは、深さ20のノードであり、平均深さは9.17であることが分かった。

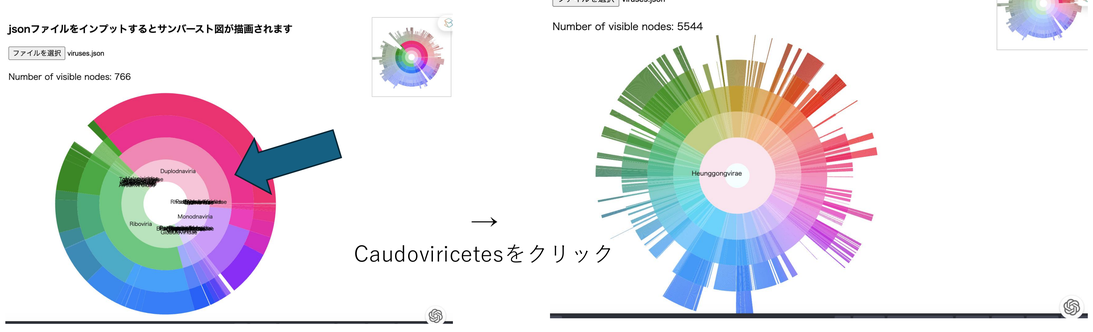
Biota
↓
Kingdom: Animalia
↓
...
↓
Class: Mammalia
↓
...
↓
Species: Homo sapiens

ヒト(Homo sapiens)
の分類樹
CoL のデータセットより

- 3つの制約の考慮
  大規模なデータセットを可視化する際は、3つの制約を考慮することが重要である[Mun14]。人間の認知能力とディスプレイの表示能力を考慮して、表示するグラフの最大5層と設定した。一方、コンピュータの計算能力の制約により、新たな課題が出た。



高々5層のみ表示する
一度に表示するノード数を減らす
全体の情報は右上のOverviewから得ることができる

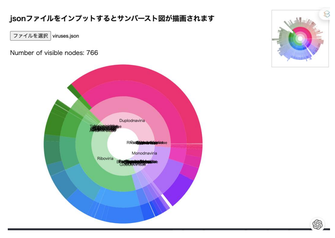


Overview

コンピュータの計算能力による課題

→ データサイズが大きい（ノード数が多い）と描画不可能

viruses.json（ノード数：1.4万、サイズ：1.2MB）
→ 問題なく描画可能



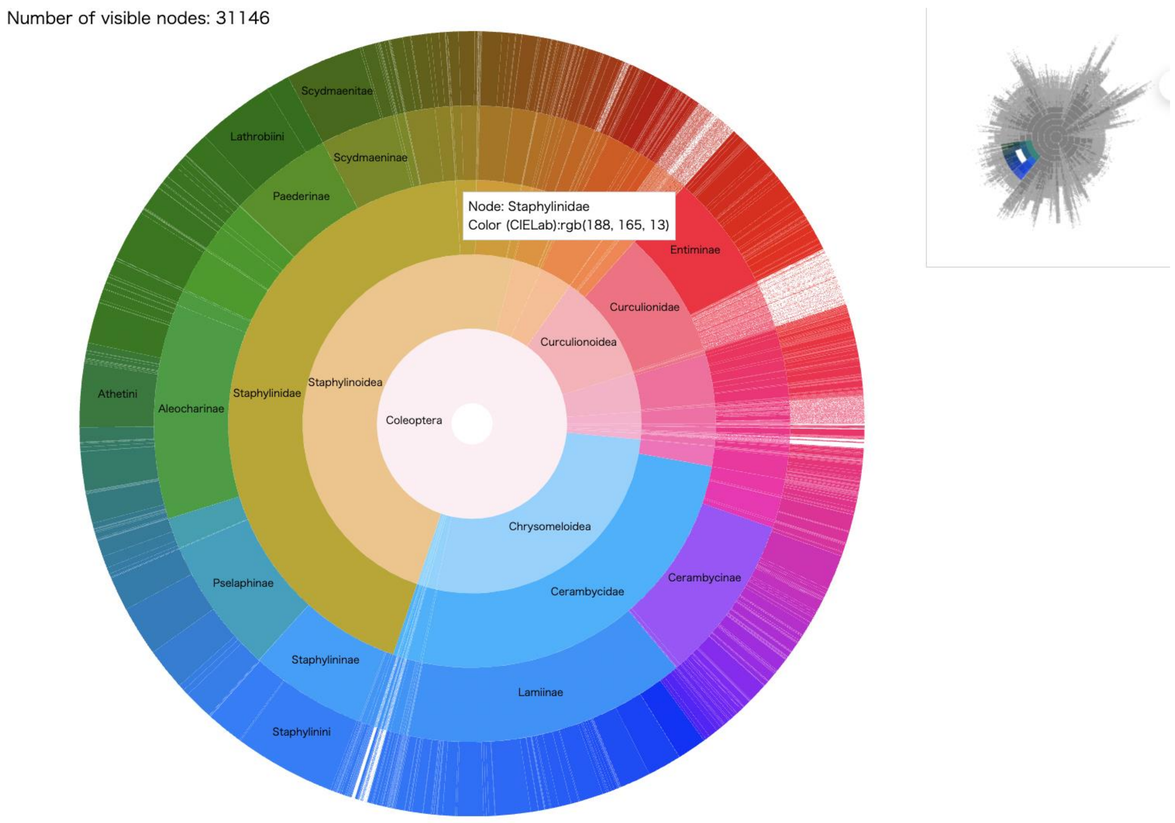
biota.json（ノード数：**200万**、サイズ：195MB）
→ 数秒経った後にエラー画面で**描画不可能**

課題への対処

本研究では、ノード数が多い可視化において、クリックのコールバック処理がパフォーマンスに与える影響を軽減することを目指している。課題の原因は、メイングラフのすべてのノード、およびOverviewの全ノードにクリックのコールバックを設けていることにある。この問題を解決するためには、クリックのコールバック数を減らすことが重要である。
具体的には、svg 要素として設定するノードおよびクリック機能やホバー機能を持つノードを限定する必要がある。
具体的に以下のように設定した。

- Overviewは、svg ではなくcanvas要素する
- ノードのサイズ（$r \times \theta$）によってクリック機能やホバー可能なノード、ラベルを表示するノードを限定する
  - $r \times \theta \geq 10$ でクリックが可能、ホバーが反応する
  - $r \times \theta \geq 50$ でラベルを表示

結果、改善前は表示不可能だった biota.json の方も時間は数秒かかるが表示することができた。（下図）



Flaskを用いた部分的データ転送

従来のアプローチでは、ユーザーがクリックするたびに全データを再送信し、再描画するため、ネットワークの負荷が増大し、反応時間が遅くなる傾向があった。しかし、大規模な木構造の可視化においては、全データの再送信は非効率的である。そこで、クリックイベントが発生した際に、Flaskを用いて必要最小限のデータのみをサーバーからクライアントに転送する方法を導入した。
具体的には、Flaskのバックエンドでクリックされたノードに関連する部分木のみを抽出し、そのデータをJSON形式でクライアントに送信する。クライアント側では、この部分データを用いて既存の可視化を部分的に更新するため、全体の再レンダリングを避けることができる。この部分的データ転送により、ネットワーク帯域の使用を最小限に抑え、クライアント側での処理負荷を軽減することが可能となる。

評価

本研究では、提案手法に基づき、200万ノードを含む大規模な木構造を正確に描画することに成功した。これにより、提案した手法が大規模データセットに対しても有効であることが確認された。
しかしながら、Flaskを用いた部分的データ転送の導入にもかかわらず、クリック後の遷移時間の短縮は実現できなかった。この結果から、現在のシステム構成やデータ処理方法にはさらなる最適化が必要であることが示唆される。

今後の展望

今後の展望として、以下の点についてさらなる研究と開発を進めていく。

■ 計算速度が遅い原因の解明
- ベンチマークを実施し、ボトルネックを特定・改善
- 角度が小さい複数の連続したノードをグラデーションのあるcanvasとして省略表示し、描画処理の効率化を図る

■ 系統樹データ以外の大規模木構造データの探索と適用
■ 他デバイス（スマートフォン、スマートウォッチ、VRなど）への適用を考慮したインターフェース設計

まとめ

本研究では、大規模で複雑な木構造データの可視化における課題に対し、色とレイアウトを統合した新しいアプローチを提案した。インタラクティブなズーム機能を導入し、ユーザーが必要な情報に焦点を当てやすくし、D3.jsを活用した柔軟な可視化を実現した。また、Flaskを用いた部分的データ転送も合わせて、200万ノードを含む大規模な木構造の正確な描画を実現し、提案手法の有効性を確認した。しかし、クリック後の遷移時間短縮には限界があり、さらなる最適化が必要であることも示唆された。今後は、さらなるシステムの最適化や他デバイスへの適用を進め、より効果的な情報可視化手法の開発を目指す。

参考文献

[Mun14] Munzner, Tamara. Visualization: Analysis and Design Chapter 1: What's Vis, and Why Do It?
[COL] Bánki, O., Roskov, Y., et al. (2024). Catalogue of Life, Amsterdam, Netherlands.
https://doi.org/10.48580/dgdwl